

# NLP HW1

## 1. Describing your methods in detail.(50%)

我的做法是先從每個句子裡面取出一個動詞和 `is, are, was` 等等。接著再各自以這些動詞為 `ROOT`，依照他的 `subtree` 由前而後依序去找主詞 (`nsubj, nsubjpass`)，把該句子的每一個主詞都走過一遍，使用兩個變數 `subj, verb` 隨時記錄目前的主詞、動詞為何。

在找到主詞、動詞之後，就繼續從該動詞的 `dependency` 去找主詞後面的受詞。值得注意的地方是這裡的受詞一定要位於主詞後面，並且存在於動詞的 `subtree` 中。受詞的找法就是從最基本的 `dobj, prep → pobj` 等等去做搜尋。每當找到一個受詞，就代表現在的變數同時包含了組 `subj, verb, obj`，就把這一組詞彙存入用來記錄的 `list` 裡面。

由上述方法找到此標準下所有可能的配對之後，就把這每一組可能的 `S, V, O` 和 `dataset.csv` 裡面的找出 `S, V, O` 去做比對，看有沒有至少一組是配對成功的。如果有，就輸出 1。如果沒有，就輸出 0。

## 2. Is there any difference between your expectations and the results? Why?(20%)

一開始我只從每一段文字裡面找出一組 `S, V, O`，拿這個 `S, V, O` 去算結果，竟然也過了 `baseline`，那時候還沒發現問題，只覺得怎麼跟預期的不太一樣，一段文字好像有很多個解，是不是 `label` 錯了。







後來轉念一想，高中的文法好像有教，`Although S. + V. + O., S. + V. + O.` 裡面其實是有兩個動詞的，因此 `label` 應該是只挑其中一組來標註。想通之後，就用暴力解，依照 `dependency` 把所有符合規則的都挑出來就成功過 `strong baseline` 了。

## 3. What difficulties did you encounter in this assignment? How did you solve it?(30%)

一開始我沒有注意到在 `string` 裡面，`he` 是包含在 `the` 的，所以 `example_with_answer.csv` 的其中一筆 `prediction` 就是錯的。發現之後，就把原來的 `string` 用空白分開，各自存在 `list` 就能解決這個問題。

另外，最一開始沒有使用 `subtree`，而是直接從每個動詞前後抓最近的主詞、受詞，表現不太好，考慮到 `dependency` 之後，就能把正確率提升到

73.5%以上了。

#	Team	Members	Score	Entries	Last	Code
1	109550074		0.75594	14	8d	
2	0816124		0.74484	6	7d	
3	0816066		0.73534	3	8d	
4	<b>0816036</b>		0.73534	1	6d	
 Your First Entry! Welcome to the leaderboard!						
5	109550068		0.72266	9	13d	